W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 24, 2020 – 11:15 am BST

PDB ID	:	4KA9
Title	:	Crystal structure analysis of single amino acid deletion mutations in EGFP
Authors	:	Arpino, J.A.J.; Rizkallah, P.J.
Deposited on	:	2013-04-22
Resolution	:	1.58 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.58 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.
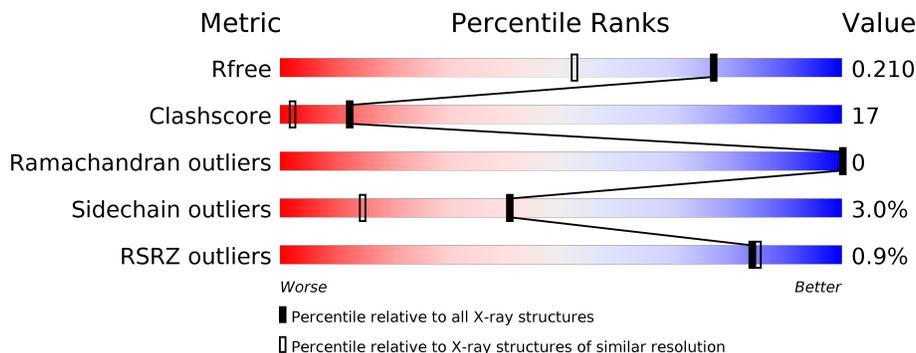
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5534 (1.60-1.56)
Clashscore	141614	5861 (1.60-1.56)
Ramachandran outliers	138981	5708 (1.60-1.56)
Sidechain outliers	138945	5703 (1.60-1.56)
RSRZ outliers	127900	5431 (1.60-1.56)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	236	77% 18%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	EDO	A	306	-	-	X	-
3	EPE	A	310	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 2295 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Green fluorescent protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	230	Total 1985	C 1259	N 335	O 384	S 7	0	19	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	VAL	-	INSERTION	UNP P42212
A	?	-	GLY	DELETION	UNP P42212
A	64	LEU	PHE	ENGINEERED MUTATION	UNP P42212
A	66	CRO	SER	CHROMOPHORE	UNP P42212
A	66	CRO	TYR	CHROMOPHORE	UNP P42212
A	66	CRO	GLY	CHROMOPHORE	UNP P42212
A	231	LEU	HIS	CONFLICT	UNP P42212

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).
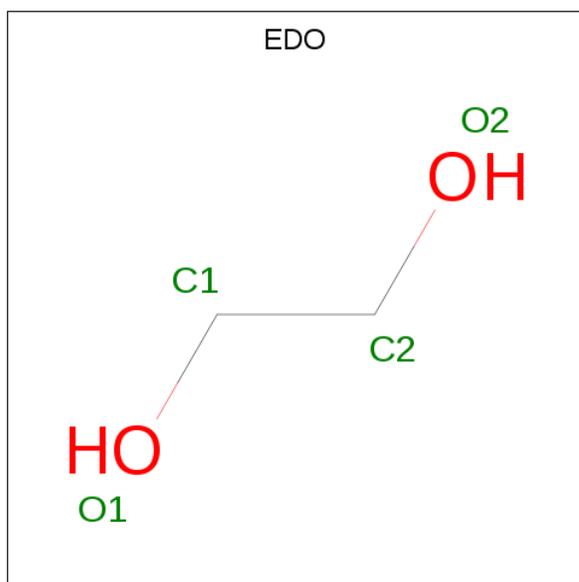
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0
2	A	1	Total 4	C 2	O 2	0	0

- Molecule 3 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).
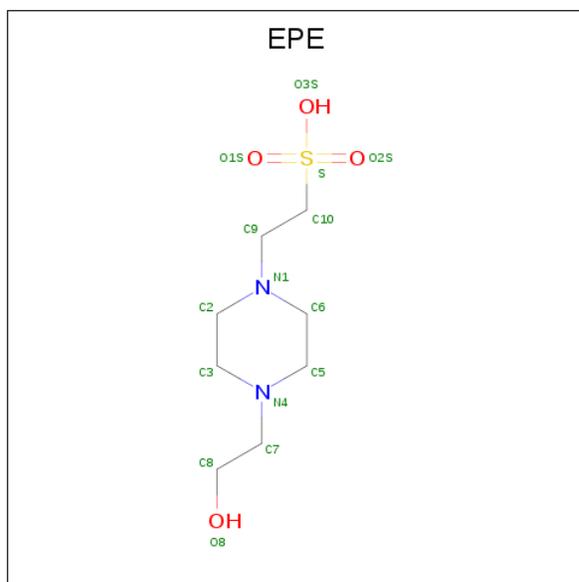
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total 15	C 8	N 2	O 4	S 1	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total 2	Na 2	0	0

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total 2	Zn 2	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	255	Total 255	O 255	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: Green fluorescent protein
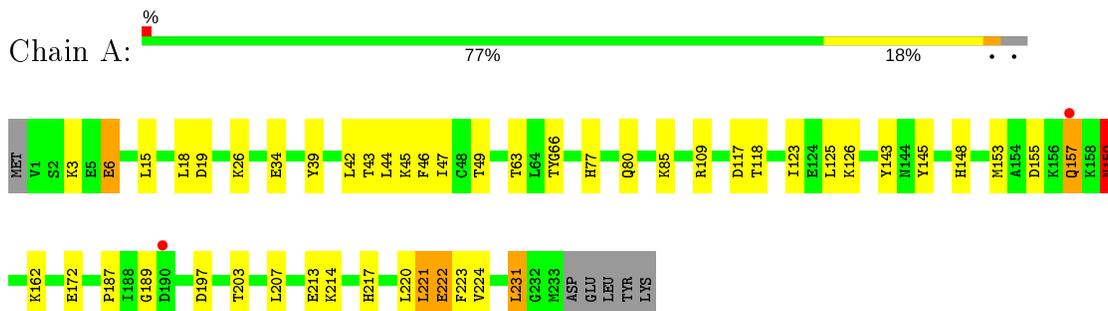
4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	91.87Å 66.66Å 45.29Å 90.00° 108.76° 90.00°	Depositor
Resolution (Å)	42.88 – 1.58	Depositor
	42.88 – 1.58	EDS
% Data completeness	97.4 (42.88-1.58)	Depositor
(in resolution range)	97.4 (42.88-1.58)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.51 (at 1.58Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.173 , 0.210	Depositor
	0.172 , 0.210	DCC
R_{free} test set	1741 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	14.5	Xtriage
Anisotropy	0.060	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 51.2	EDS
L-test for twinning[2]	$<|L|> = 0.46$, $<L^2> = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2295	wwPDB-VP
Average B, all atoms (Å^2)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.03% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, ZN, EPE, EDO, CRO

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	1.22	4/2003 (0.2%)	1.20	5/2705 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	39	TYR	CB-CG	-5.69	1.43	1.51
1	A	222	GLU	CD-OE2	5.51	1.31	1.25
1	A	189	GLY	N-CA	5.47	1.54	1.46
1	A	143	TYR	CE1-CZ	5.25	1.45	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	197	ASP	CB-CG-OD2	-5.72	113.15	118.30
1	A	19	ASP	CB-CG-OD1	5.49	123.24	118.30
1	A	117	ASP	CB-CG-OD1	5.08	122.88	118.30
1	A	221[A]	LEU	CB-CG-CD1	5.02	119.53	111.00
1	A	221[B]	LEU	CB-CG-CD1	5.02	119.53	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	159[B]	ASN	Peptide

5.2   Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1985	0	1932	62	0
2	A	36	0	54	8	1
3	A	15	0	18	15	0
4	A	2	0	0	1	0
5	A	2	0	0	0	0
6	A	255	0	0	20	3
All	All	2295	0	2004	69	4

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 17.

The worst 5 of 69 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:221[B]:LEU:HD21	1:A:223:PHE:HE2	1.09	1.12
3:A:310:EPE:H92	6:A:624:HOH:O	1.65	0.97
2:A:307:EDO:H12	6:A:633:HOH:O	1.62	0.96
1:A:221[B]:LEU:HD21	1:A:223:PHE:CE2	2.00	0.95
1:A:172[B]:GLU:HG3	6:A:532:HOH:O	1.66	0.94

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
6:A:571:HOH:O	6:A:571:HOH:O[2_555]	1.35	0.85
6:A:617:HOH:O	6:A:618:HOH:O[2_555]	1.87	0.33
2:A:304:EDO:C2	2:A:306:EDO:O2[4_545]	2.10	0.10
6:A:587:HOH:O	6:A:587:HOH:O[2_554]	2.14	0.06

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	244/236 (103%)	238 (98%)	6 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	219/207 (106%)	212 (97%)	7 (3%)	39	13

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	159[A]	ASN
1	A	231	LEU
1	A	159[B]	ASN
1	A	145	TYR
1	A	214	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	157	GLN
1	A	184	GLN
1	A	212	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
1	CRO	A	66	1	23,23,24	2.29	10 (43%)	30,32,34	2.05	7 (23%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	CRO	A	66	1	-	0/12/31/32	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	66	CRO	CA2-C2	-5.86	1.42	1.48
1	A	66	CRO	CB2-CA2	5.63	1.39	1.35
1	A	66	CRO	C2-N3	-2.81	1.33	1.39
1	A	66	CRO	CE1-CD1	2.45	1.43	1.38
1	A	66	CRO	C1-N3	-2.42	1.33	1.37

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	66	CRO	CA2-C2-N3	7.17	106.76	103.37
1	A	66	CRO	C1-CA1-N1	-3.52	104.25	109.96
1	A	66	CRO	O2-C2-CA2	-3.01	129.27	130.96
1	A	66	CRO	CB2-CA2-C2	2.63	125.41	122.28

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	CRO	O3-C3-CA3	-2.57	118.62	126.39

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	66	CRO	1	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 14 ligands modelled in this entry, 4 are monoatomic - leaving 10 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
2	EDO	A	304	-	3,3,3	0.21	0	2,2,2	1.47	1 (50%)
2	EDO	A	308	-	3,3,3	0.22	0	2,2,2	1.35	0
3	EPE	A	310	-	15,15,15	2.07	2 (13%)	18,20,20	2.81	11 (61%)
2	EDO	A	301	-	3,3,3	0.25	0	2,2,2	0.49	0
2	EDO	A	305	-	3,3,3	0.49	0	2,2,2	0.65	0
2	EDO	A	302	-	3,3,3	1.28	0	2,2,2	0.65	0
2	EDO	A	309	-	3,3,3	0.98	0	2,2,2	0.44	0
2	EDO	A	303	-	3,3,3	0.49	0	2,2,2	0.67	0
2	EDO	A	306	-	3,3,3	0.24	0	2,2,2	0.86	0
2	EDO	A	307	-	3,3,3	0.53	0	2,2,2	0.18	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	EDO	A	304	-	-	1/1/1/1	-
2	EDO	A	308	-	-	1/1/1/1	-
3	EPE	A	310	-	-	6/9/19/19	0/1/1/1
2	EDO	A	301	-	-	1/1/1/1	-
2	EDO	A	305	-	-	0/1/1/1	-
2	EDO	A	302	-	-	0/1/1/1	-
2	EDO	A	309	-	-	0/1/1/1	-
2	EDO	A	303	-	-	0/1/1/1	-
2	EDO	A	306	-	-	1/1/1/1	-
2	EDO	A	307	-	-	0/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	310	EPE	C10-S	-7.04	1.67	1.77
3	A	310	EPE	O1S-S	2.82	1.53	1.45

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	310	EPE	C5-N4-C3	5.69	121.65	108.83
3	A	310	EPE	C9-N1-C2	4.09	121.69	111.23
3	A	310	EPE	C6-N1-C2	3.85	117.50	108.83
3	A	310	EPE	C9-N1-C6	3.74	120.80	111.23
3	A	310	EPE	C7-N4-C5	3.50	120.20	111.23

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	310	EPE	C10-C9-N1-C2
3	A	310	EPE	S-C10-C9-N1
3	A	310	EPE	C9-C10-S-O2S
3	A	310	EPE	C9-C10-S-O3S
2	A	304	EDO	O1-C1-C2-O2

There are no ring outliers.

6 monomers are involved in 24 short contacts:

wwPDB X-ray Structure Validation Summary Report

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	304	EDO	0	1
3	A	310	EPE	15	0
2	A	305	EDO	1	0
2	A	303	EDO	1	0
2	A	306	EDO	3	1
2	A	307	EDO	3	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	229/236 (97%)	-0.08	2 (0%)	84	85	9, 13, 24, 48	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	157	GLN	2.8
1	A	190[A]	ASP	2.5

6.2 Non-standard residues in protein, DNA, RNA chains (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
1	CRO	A	66	22/23	0.97	0.07	10,11,13,17	0

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

wwPDB X-ray Structure Validation Summary Report

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	A	307	4/4	0.71	0.28	54,56,58,58	0
3	EPE	A	310	15/15	0.74	0.24	51,53,70,75	0
2	EDO	A	308	4/4	0.82	0.16	29,35,39,40	0
2	EDO	A	302	4/4	0.85	0.14	24,29,29,31	0
2	EDO	A	306	4/4	0.92	0.16	28,28,33,34	0
2	EDO	A	305	4/4	0.93	0.12	30,32,32,33	0
2	EDO	A	303	4/4	0.95	0.09	25,27,34,34	0
2	EDO	A	309	4/4	0.96	0.11	15,16,20,21	0
2	EDO	A	304	4/4	0.96	0.10	24,29,31,32	0
2	EDO	A	301	4/4	0.97	0.08	26,28,29,32	0
4	NA	A	311	1/1	0.98	0.09	16,16,16,16	0
4	NA	A	314	1/1	0.99	0.09	13,13,13,13	0
5	ZN	A	312	1/1	1.00	0.05	19,19,19,19	0
5	ZN	A	313	1/1	1.00	0.05	12,12,12,12	1

6.5 Other polymers (i)

There are no such residues in this entry.